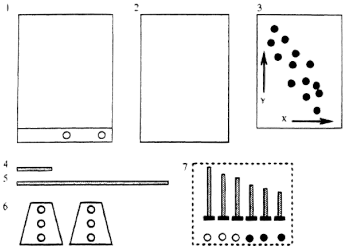
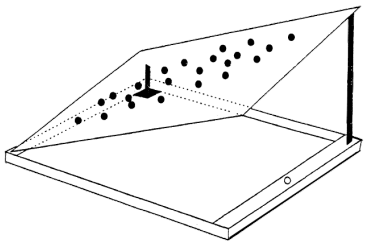
Causation laboratory background

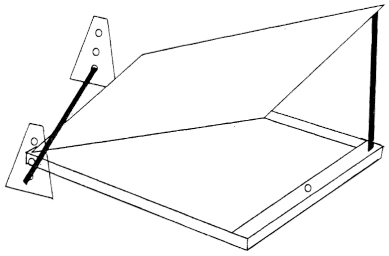
Consider the physical laboratory involving the plane.



Materials



Points on a plane



Experimental set up

The equation of the plane can be written as

$$y = \alpha + \beta x + \gamma z$$

which has no error. All of the points lie exactly on the plane and both β and γ are unknown.

In the laboratory, the value of z corresponds to the height of the plane off of the horizontal. The six tower-markers were available of varying heights. The height represents the fixed z value for that tower-marker; it is unchangeable for any tower-marker. The planes were held in place in such a way that, at least approximately, the values of the parameters are fixed, if unknown.

The value of x assigned to a tower-marker is whatever we choose which consequently will determine the value of y as determined by the plane. In this respect, setting the value of x **causes** the value of y as we move the tower-marker along the value x until the tower-marker strikes the plane to determine$y$. The sign and magnitude of β reveal the nature of the causal relationship between x and y (for fixed z). If, for example, $\beta > 0$ then for any fixed z (i.e. tower-marker) as we increase the value of x, the value of y will also increase accordingly. If $\beta < 0$, then as x increases y decreases, again for fixed z (the tower-marker's height). When $\beta = 0$, x can be changed as much as we like and y will not change (for z fixed); that is changes in x will **not cause** changes in y (for z fixed).

There will of course be some error in our measurements of x and y caused both by the act of measuring and by the steadiness with which the plane has been held.

Data have been collected on many (x, y) pairs measured by several teams, each of which were following three different experimental plans: `"observational"`, `"randomized"`, or `"randomizedBlock"`. Only the last of these were based on knowing something about the possible value of z (being on of small, medium, or large). The data appears in the assignment data directory as the file `labData.csv`.

The purpose of the laboratory is to investigate and compare the different experimental plans. Of particular interest is whether or not x causes y, that is $H_0 : \beta = 0$. And if so, to estimate the value of β defining the causal relationship.

In practice, the values of z are never known – there is always some variate that is not measured, perhaps not even thought of, that might be part of relating y to x. These are called **lurking variables** and they will always exist.

There were 18 teams of students (2 per team) who carried out the studies. Each team reported their measurements of x and y of the tower-markers with x assigned and y determined by when the tower-marker struck the plane for that x.

The first protocol was denoted "`observational`" and coincided with the tower-markers from shortest to tallest being assigned to six different x values, namely 5, 10, 15, 20, 25, and 30.

The second protocol was denoted "`randomized`" and randomly assigned half of the tower-markers to $x = 5$ and the remaining half to $x = 25$. Each team conducted two replications (`rep`) of this assignment and measured y values for each.

The third protocol, denoted "`randomizedBlock`", first paired the tower-markers according to height to produce three pairs (or blocks): the two shortest, the two tallest, and the remaining two of middling height. One tower-marker in each pair was randomly assigned to $x = 5$, the other to $x = 25$. Again, each team conducted two replications (`rep`) of this assignment and measured y values for each.

The first few rows of `labData` are

```
##   team          type rep  x    y
## 1    1 observational   1  5 24.5
## 2    1 observational   1 10 23.0
## 3    1 observational   1 15 21.0
## 4    1 observational   1 20 20.0
## 5    1 observational   1 25 18.0
## 6    1 observational   1 30 15.0
```

And the last few are:

```
##     team            type rep  x  y
## 535   18 randomizedBlock   2  5  6
## 536   18 randomizedBlock   2  5 11
## 537   18 randomizedBlock   2  5 20
## 538   18 randomizedBlock   2 25 11
## 539   18 randomizedBlock   2 25 26
## 540   18 randomizedBlock   2 25 35
```

Summary statistics are

```
##       team                    type          rep            x
##  Min.   : 1.0   observational  :108   Min.   :1.0   Min.   : 5.0
##  1st Qu.: 5.0   randomized     :216   1st Qu.:1.0   1st Qu.: 5.0
##  Median : 9.5   randomizedBlock:216   Median :1.0   Median :17.5
##  Mean   : 9.5                         Mean   :1.4   Mean   :15.5
##  3rd Qu.:14.0                         3rd Qu.:2.0   3rd Qu.:25.0
##  Max.   :18.0                         Max.   :2.0   Max.   :30.0
##        y
##  Min.   : 0.0
##  1st Qu.:14.0
##  Median :19.0
##  Mean   :18.8
##  3rd Qu.:24.0
##  Max.   :37.0
```